FRET Study of Membrane Proteins: Simulation-Based Fitting for Analysis of Membrane Protein Embedment and Association

Petr V. Nazarov,*† Rob B. M. Koehorst,* Werner L. Vos,* Vladimir V. Apanasovich,† and Marcus A. Hemminga*
*Laboratory of Biophysics, Wageningen University, Wageningen, The Netherlands; and †Department of Systems Analysis, Faculty of Radio Physics, Belarusian State University, Minsk, Belarus

ABSTRACT A new formalism for the simultaneous determination of the membrane embedment and aggregation of membrane proteins is developed. This method is based on steady-state Förster (or fluorescence) resonance energy transfer (FRET) experiments on site-directed fluorescence labeled proteins in combination with global data analysis utilizing simulation-based fitting. The simulation of FRET was validated by a comparison with a known analytical solution for energy transfer in idealized membrane systems. The applicability of the simulation-based fitting approach was verified on simulated FRET data and then applied to determine the structural properties of the well-known major coat protein from bacteriophage M13 reconstituted into unilamellar DOPC/DOPG (4:1 mol/mol) vesicles. For our purpose, the cysteine mutants Y24C, G38C, and T46C of this protein were produced and specifically labeled with the fluorescence label AEDANS. The energy transfer data from the natural tryptophan at position 26, which is used as a donor, to AEDANS were analyzed assuming a helix model for the transmembrane domain of the protein. As a result of the FRET data analysis, the topology and bilayer embedment of this domain were quantitatively characterized. The resulting tilt of the transmembrane helix of the protein is 18 ± 2°. The tryptophan is located at a distance of 8.5 ± 0.5 Å from the membrane center. No specific aggregation of the protein was found. The methodology developed here is not limited to M13 major coat protein and can be used in principle to study the bilayer embedment of any small protein with a single transmembrane domain.

INTRODUCTION

Membrane proteins play an important role in almost all cell activities. They perform a staggering range of biological reactions including respiration, signal transfer, and molecular and ion transport (1). However, the structure determination of membrane proteins is still at the frontier of structural biology. Although 30–40% of all proteins are membrane proteins, <1% of the known protein structures are for membrane proteins (2,3). (For the most recent state for membrane proteins of known structure, see: http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html.) The complexity and delicacy of membrane-protein systems substantially impede the application of standard methods of protein study, such as x-ray crystallography and NMR (3,4). Furthermore, these techniques are aimed at short-range structural information, and do not seem to be useful for the study of long-range interactions; for instance, in the case of protein association and clustering.

These factors impel to find other approaches to study proteins incorporated into lipid bilayers. A successful alternative is Förster (or fluorescence) resonance energy transfer (FRET) spectroscopy (5–7). This technique provides distance information within a range of 10–100 Å, which is sufficient to study the structure of membrane proteins and their complexes. FRET spectroscopy has been successfully applied to several problems in biology as a means of estimating intra- and intermolecular distances in macromolecular systems, especially proteins (7–9). The idea of using FRET is in the labeling of the macromolecules with fluorescent labels of two kinds—a donor and an acceptor, and analysis of radiationless dipole-dipole energy transfer (10) between them. One of the advantages of such an approach is that several natural amino acid residues of a protein, such as Trp and Tyr, can be utilized as fluorescent labels (4,11).

Despite the elegant analytical models for a uniform planar donor-acceptor distribution that were developed two decades ago (12–14), the complexity of protein-lipid systems hampers and limits an analytical interpretation of FRET data (15,16). For example, in the present work several numerical tests were performed to study the applicability of analytical models for the analysis of membrane protein systems. It was found that analytical expressions give incorrect results when the size of acceptor-host molecules is comparable to the Förster distance of the donor-acceptor pair. On the other hand, simulation modeling of photophysical processes in an experimental system during a fluorescence measurement was proven to be a powerful alternative to analytical modeling, not restricted to special conditions (15–18). The standard approaches to simulate FRET effects in complex systems are various Monte Carlo simulation schemes (15–17). However, Monte Carlo simulation modeling is a very time-consuming operation. Furthermore, a time-resolved approach is not needed for the analysis of steady-state FRET data.

The goal of the current work is to develop and test a methodology for the analysis of steady-state FRET data to build a low-resolution structural model of a protein-membrane system with a quantitative characterization of its parameters. To perform this goal, a steady-state FRET model

Submitted February 6, 2006, and accepted for publication April 7, 2006.

Address reprint requests to M. A. Hemminga, Tel.: 31-317-482635 or 31-317-482044; Fax: 31-317-482725; E-mail: marcus.hemminga@wur.nl.

0006-3495/06/07/454/13 $2.00
doi: 10.1529/biophysj.106.082867

is built and utilized in a simulation-based fitting (SBF) approach to approximate the experimental data by their simulated analogs (17,18). By comparison with standard analytical data-fitting techniques, simulation modeling has the advantage that it operates with the physical parameters of the system itself and gives a direct insight into how they affect the experimental characteristics of the system.

The methodology developed is tested on a well-known coat protein from bacteriophage M13. During a part of its life cycle, the coat protein is stored as a membrane protein in the *Escherichia coli* host. Therefore, it is an excellent model system to study fundamental aspects of protein-lipid and protein-protein interactions (19). This single membrane-spanning protein consists of 50 amino-acid residues and has mainly an α-helical conformation. The protein has been extensively studied in model membrane systems by several biophysical techniques (19–30). For FRET studies, the natural single tryptophan residue of the protein at position 26 (Trp-26) was used as a donor label. To introduce an acceptor label to the protein, a number of mutants, containing unique cysteine residues at specific positions, were produced. The cysteine residues were labeled with the fluorescent label *n*-(acetylaminoethyl)-5-naphthylamine-1-sulfonic acid (AEDANS) (20). This label was used as an acceptor. To separate intra- and intermolecular energy transfer contributions, we performed titration experiments in which we added wild-type protein to mutant proteins at different L/P ratios. Both unlabeled mutant and wild-type protein can be considered spectroscopically identical as donor-containing molecules without acceptor label. The labeled mutants contain both a donor and acceptor. In a fluorescence excitation experiment one can optically select the labeled mutant proteins by monitoring the acceptor fluorescence. In a fluorescence excitation spectrum FRET can be deduced from the enhancement of acceptor fluorescence at the donor absorption wavelength. Upon addition of donor containing wild-type protein, the intermolecular energy transfer component is increased exclusively.

In this article we focus our analysis on the transmembrane domain of the protein, which was recently found to be in an almost perfect α-helix conformation (25,30). To take into account the membrane embedment of the proteins and possible protein aggregation, a model of a protein-lipid bilayer system is generated. This model is then used in an SBF approach to analyze the fluorescence data. To make the SBF procedure more effective, a global analysis strategy is applied, in which all data are analyzed simultaneously. This approach provides information about the membrane embedment of the transmembrane protein domain in terms of protein depth, tilt angle, and protein association.

EXPERIMENTAL

Sample preparation

The lipid bilayer systems were prepared from dioleoylphosphatidylcholine (DOPC, 18:1PC) and dioleoylphosphatidylglycerol (DOPG) lipids in a 4:1 molar ratio, denoted as DOPC/DOPG. DOPC was purchased from Avanti Polar Lipids (Alabaster, AR) and DOPG was purchased from Sigma (St. Louis, MO).

Site-specific cysteine mutants of M13 major coat protein were prepared, purified, and labeled with AEDANS (Molecular Probes, Eugene, OR) as described previously (20). Wild-type protein and AEDANS-labeled M13 coat protein mutants were reconstituted into phospholipid bilayers as reported earlier (31).

For this study we used AEDANS-labeled cysteine mutants of M13 coat protein with the cysteine residue at positions 24 (Y24C), 38 (G38C), and 46 (T46C). Titration experiments were performed in which the wild-type protein concentration was increased, whereas the mutant concentration was kept constant. The sample conditions for these titrations are given in Table 1. For the purpose of correcting the fluorescence results (see $\varepsilon_A^{290}/\varepsilon_A^{340}$ in Eq. 1), we also used a mutant (Y21A/Y24A/W26A/G23C) having the AEDANS-labeled cysteine at position 23, in combination with a threefold mutation of the tryptophan at position 26 and the tyrosines at positions 21 and 24 into alanines. The labeling efficiency of the mutants having the AEDANS label at position 24, 38, and 46 was determined as reported previously (32) and amounted to 62, 55, and 69%, respectively. The labeling efficiency is explicitly taken into account in Table 1 in the ratio of the number of unlabeled to labeled proteins (r_{ul}), as it affects the acceptor concentration and therefore the energy transfer efficiency.

For the fluorescence experiments, stock solutions of protein mutants and wild-type protein solubilized in cholate buffer were mixed with solutions of lipids in the same buffer, as described previously (31). Repeated dialysis of the

TABLE 1 Sample composition of M13 major coat protein incorporated into DOPC/DOPG bilayers given in terms of r_{LP} and r_{ul}, and observed energy transfer efficiencies E for mutants with acceptor positions n_A at 24, 38, and 46

Data set	1	2	3	4
n_A	24	38	38	46
r_{LP}	3600	209	3213	105
r_{ul}	0.6	6	1	1.3
E	0.558	0.121	0.254	0.152
r_{LP}	1059	128	553	80
r_{ul}	4.5	10	10	2.2
E	0.165	0.094	0.056	0.147
r_{LP}	621	71	303	55
r_{ul}	8.4	19	18	3.9
E	0.099	0.071	0.043	0.135
r_{LP}	340	45	159	38
r_{ul}	16	33	36	6
E	0.058	0.056	0.027	0.127
r_{LP}	179	28	65	25
r_{ul}	32	54	88	10.4
E	0.033	0.047	0.020	0.116

For mutant G38C, two FRET titration experiments were carried out at different values of r_{LP} and r_{ul}.

mixtures in cholate-free buffer was performed to remove the cholate in the sample. The lipid loss during dialysis can vary near 20% (31), and this fact should be taken into account during the analysis of the experimental data.

Fret experiments

Optical spectroscopy

Absorption spectra were recorded on a Varian Cary 5E UV-Vis-NIR spectrophotometer (Cary, NC) and fluorescence emission and fluorescence excitation measurements were performed on a Fluorolog 3.22 manufactured by Jobin Yvon-Spex (Edison, NJ) as described elsewhere (30,33). For fluorescence excitation measurements, the detection wavelength was set at the maximum of the acceptor (AEDANS) fluorescence of a particular mutant and the excitation wavelength was scanned from 260 to 400 nm. The detection wavelength was different for each mutant, because the AEDANS fluorescence maximum varies with bilayer depth (i.e., local polarity) of the AEDANS label and therefore with the residue number of the labeled cysteine. The AEDANS fluorescence for mutants 24 and 46 was observed at 490 nm; for mutant 38, this was 470 nm. The applied slit-widths of the detection and excitation monochromators corresponded to 5- and 2-nm bandpass, respectively. The spectra were automatically corrected on the Fluorolog 3.22 for variations in the lamp output by dividing the sample signal by that of an internal reference detection system. All excitation spectra were corrected for background fluorescence using an equimolar solution of pure wild-type protein (no AEDANS present). Moreover, tryptophan fluorescence is neglectable at the detection wavelength (see Fig. 4 *A*), therefore the observed radiation exclusively belongs to AEDANS. The temperature during all measurements was 20°C. Because of the small protein concentrations used in our experiments ($\sim$1 μM), errors caused by the inner filter effects can be neglected.

Analysis of AEDANS excitation spectra

The derivation of the mathematical expressions for the analysis of the experimental excitation spectra is given in Appendix A. For our analysis, we used the energy transfer efficiency E, which can be calculated from the fluorescence intensities (8) by

$$E = \frac{1}{1 + r_{ul}}\left(\frac{F^{290}}{F^{340}} - \frac{\varepsilon_A^{290}}{\varepsilon_A^{340}}\right)\frac{\varepsilon_A^{340}}{\varepsilon_D^{290}}, \quad (1)$$

where r_{ul} is the ratio of the number of unlabeled to labeled proteins. For every sample, the ratio of the fluorescence intensity at 290 nm, F^{290} (mainly donor excitation) to that at 340 nm, F^{340} (exclusively acceptor excitation) was calculated, being a measure of the donor-to-acceptor energy transfer. The ratio F^{290}/F^{340} was corrected for direct excitation of AEDANS at 290 nm by subtracting the ratio of the extinction coefficients $\varepsilon_A^{290}/\varepsilon_A^{340} = 0.20$ (this ratio was calculated using mutant Y21A/Y24A/W26A/G23C). Finally, the ratio of the extinction coefficients of the acceptor at 340 nm (ε_A^{340}) and donor at 290 nm (ε_D^{290}) have to be taken into account in Eq. 1 ($\varepsilon_A^{340}/\varepsilon_D^{290} = 1.2$).

METHODOLOGY

Model for the transmembrane domain of M13 coat protein incorporated into a lipid bilayer

The proposed simplified structural model for the transmembrane domain of M13 coat protein consists of an ideal α-helix (Fig. 1) (19,25–27). The complete set of structural parameters that determines the protein-lipid system is presented in Table 2. In the protein model, we assume two specific sites: a donor and an acceptor site that will enable us to calculate the theoretical energy transfer and relate that to the FRET experiments. For the M13 coat protein, which consists of 50 amino-acid residues, the donor is the Trp-26 and the acceptor is introduced at an arbitrary position in the

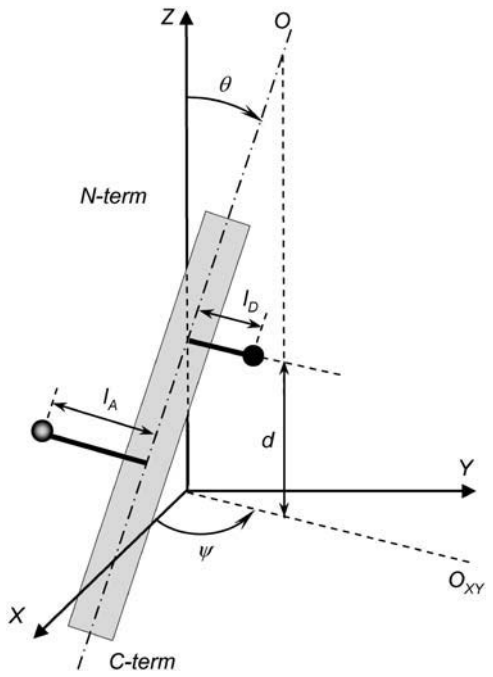

FIGURE 1 (*A*) Schematic drawing of the transmembrane domain of M13 major coat protein consisting of an ideal α-helix (19,25–27). As an example, the donor (Trp-26, *solid circle*, located on the N-terminal side at a distance l_D from the protein helix axis) and acceptor (AEDANS, *shaded circle*, located on the C-terminal side at a distance l_A from the protein helix axis) are attached at positions 26 and 38, respectively. The membrane axis system is indicated by *X*, *Y*, and *Z*. The *XY* plane at $Z = 0$ corresponds to the center of the lipid bilayer in which the protein is inserted. Parameter d is the distance from the origin of the coordinate system of the protein to the center of the lipid bilayer. Axis *O* is the helix axis of the protein domain, and θ is the tilt angle, i.e., the angle between the helix axis and the normal to the membrane. The value O_{xy} is the projection of the helix axis on the *XY* plane. Angle ψ is the protein tilt direction, i.e., the direction of the tilting of the helix. The complete set of structural parameters that determines the protein-lipid system is presented in Table 2.

TABLE 2 Definition of the parameters used in the model for the protein-lipid system

Parameter	Range/value	Unit	Description
n_0	26	—	The position of a reference amino-acid residue. The projection of its C_α to the helix axis of the protein O gives the origin of the coordinate system of the protein. Position $n_0 = 26$ was selected for the transmembrane domain of M13 major coat protein.
H	1.5	Å	Translation per amino-acid residue along the helix; this is 1.5 Å for a perfect α-helix.
n_r	3.6	—	Number of amino-acid residues per one turn; this is 3.6 for a perfect α-helix.
n_D	26	—	Donor position; position of amino-acid residue given by the donor. For M13 coat protein, the donor is Trp-26, which is located in the transmembrane domain.
n_A	1 – 50	—	Acceptor position; position of amino-acid residue labeled by the acceptor. For the transmembrane domain of M13 coat protein, the acceptor positions are 24, 38 and 46.
$_D$	6.5	Å	Donor arm, the average distance from the donor moiety to the helix axis. A value $l_D = 6.5$ Å was taken (25).
l_A	9.5	Å	Acceptor arm, the average distance from the acceptor moiety to the helix axis. A value $l_A = 9.5$ Å was taken (25).
θ	0 – 90	°	Protein tilt angle; the angle between the helix axis and the normal to the membrane.
D	0 – 30	Å	Distance from the origin of the coordinate system of the protein to the center of the bilayer.
ψ	−180 – 180	°	Protein tilt direction; the direction of the protein transmembrane domain tilting. A value $\psi = 0$ means that protein is tilted toward the C_α of the reference (n_0) amino-acid residue.
N_P	500	—	Number of proteins in the system. All simulations were performed for models containing 500 proteins.
S_L	72	Å^2	Area occupied by a lipid in one leaflet of a bilayer; the average area for the DOPC/DOPG system is 72 Å^2 (24).
L	0.0 – 1.0	—	Lipid loss; ratio of lipids lost during dialysis to their initial quantity.
D_P	10	Å	Protein exclusion distance; minimal protein-protein distance. For M13 coat protein, a value $D_P = 10$ Å was taken.
r_{LP}	≥0	—	Lipid/protein ratio.
r_{ul}	≥0	—	Ratio between the number of unlabeled and labeled proteins.
k	0.0–1.0	—	Protein-protein association probability, defined as the percentage of clustered proteins with respect to the total number of proteins (see Fig. 2).
R_0	24	Å	Förster distance. A value of 24 Å is calculated using data from the photophysical properties of the donor and acceptor.

In the simulations, parameters θ, d, ψ, L, and k are varied. Parameters n_A, r_{LP}, and r_{ul} are determined by the experiment; the other parameters are fixed as shown in the Table.

transmembrane protein domain via cysteine mutagenesis and labeling with a fluorescent label (in our case: AEDANS). Acceptor sites are empty for nonlabeled or wild-type proteins.

As a model for proteins incorporated into a lipid bilayer, a square region of a bilayer containing a certain number of proteins (N_P) is considered. By using a three-dimensional mathematical description, protein molecules as shown in Fig. 1 are inserted randomly (both in location as well as in orientation) into the lipid bilayer in the way that the angle θ between the membrane normal and their main axis O of the transmembrane domain is between 0 and 90°. The direction of the protein tilt is given by ψ. A value $\psi = 0$ means that protein is tilted toward the C_α of the reference (n_0) amino-acid residue. The depth of protein insertion is given by parameter d. It is assumed that, when inserted into the membrane, the proteins occupy a cylindrical region in both bilayer leaflets with a protein exclusion distance D_P. Within this region no lipids or other proteins can be located.

In the protein-lipid model, the direction of the tilt and the orientation of the N-terminal domain of each protein in the coordinate system of the bilayer are set randomly. Two algorithms of protein insertion were considered. In the first one, three reference points located in the transmembrane domain were selected, and during insertion the distances between these reference points of the inserted protein and similar points on the nearest proteins were compared with D_P to determine the overlapping situation. In case of a clash, the algorithm selected a different protein direction or, if still unsuccessful after a number of tries, a new protein position. In the second algorithm, the proteins were simply inserted randomly at distances larger than D_P. In this case, their tilted transmembrane domains could, in principle, overlap. This first algorithm turned out to be quite time-consuming (from 2 to 20 times, depending on the L/P ratio) without significant changes in the energy transfer results (<0.02 for the extreme case of L/P 25). Therefore, we decided to use the simplified algorithm in all further fitting procedures.

The area of the considered square region of the membrane is calculated from the experimental L/P ratio (r_{LP}), the protein exclusion distance (D_P), the area per two lipid molecules (S_L), and the ratio of lipids lost during dialysis to their initial quantity (i.e., the lipid loss L), as

$$S = N_P\left(S_L r_{LP}(1 - L)/2 + \pi D_P^2/4\right). \quad (2)$$

Furthermore, to be able to work with mixtures of labeled and unlabeled protein molecules, the ratio r_{ul} between the number of unlabeled and labeled proteins needed to be introduced into the model.

Similar to the experimental reconstituted protein-lipid system, protein molecules can be inserted into the model membrane randomly with parallel and antiparallel orientations; this means that the N-terminal domain of the protein can be located either in the upper or in the lower leaflet of the membrane with equal probabilities. The result of these equiprobable orientations is that the membrane system contains two layers of donors and two layers of acceptors.

A protein-protein association probability k is introduced to take into account the ability of the membrane proteins to form oligomers or clusters. The algorithm for this association is as follows. All proteins are divided into two groups: free and associated. Initially, the coordinates of the free proteins in the XY plane of the membrane are randomly generated. Before incorporation of a new protein into the membrane model, it is checked whether the position for the protein is free (all previously incorporated proteins are not closer than D_P). If the position is occupied, random coordinates are selected again. For associated proteins, the algorithm is slightly changed: the XY coordinates are selected to incorporate the protein at a distance D_P next to one of the previously incorporated proteins. The value of k ranges from 0 to 1, indicating no-association and complete-association (all proteins are clustered together), respectively. The effect of protein association is exemplified in Fig. 2.

Apart from the structural parameters and parameters related to the composition of the protein-lipid system, one additional physical parameter needs to be introduced: this is the Förster distance R_0 of the donor-acceptor pair. Its physical meaning is discussed below.

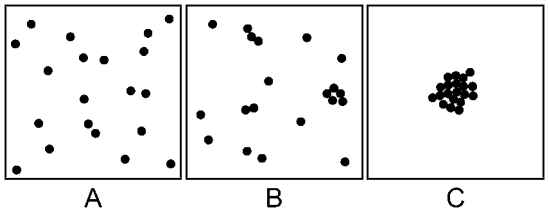

FIGURE 2 Schematic illustration of the effect of protein association resulting from the model described in the text. (*A*) Random distribution of proteins with $k = 0$, (*B*) partially associated proteins with $k = 0.5$, and (*C*) completely associated proteins ($k = 1$). Proteins are schematically indicated by solid dots. The figures show that at increasing values of k, the proteins aggregate into clusters in a nonspecific way.

Models for FRET

Basic model for energy transfer

Being in an excited state, a fluorescent molecule has a dipole-dipole interaction with other molecules in close proximity—which can lead to energy transfer from the excited molecule to the nonexcited ones. If we assume that the emission spectrum of the donor overlaps with the absorption spectrum of the acceptor, the photon absorbed by the donor can be transferred to the acceptor with a rate constant k_{ET} depending on the sixth power of the distance between the donor and acceptor,

$$k_{ET} = \frac{1}{\tau_D}\left(\frac{R_0}{R}\right)^6, \tag{3}$$

where τ_D is the lifetime of an isolated donor, and R is the distance between the donor and acceptor. The so-called Förster distance R_0 is given by

$$R_0 = 9780(\kappa^2 n^{-4} Q_D J)^{1/6}. \tag{4}$$

In this equation, κ^2 is the orientation factor describing the relative orientation of the transition dipole moments of the donor and the acceptor, n is the refractive index of the environment, Q_D is the quantum yield of an isolated donor, and J is the integral expressing the degree of donor emission and acceptor absorption spectral overlap (5).

Consider now a system of multiple donors and acceptors that are fixed at their positions. Let us number the donors $i = 1\ldots N_D$, and acceptors $j = 1\ldots N_A$. Here, N_D is the number of donor molecules and N_A is the number of acceptor molecules. The probability for each donor to transfer energy to one of the acceptors can then be calculated as

$$p_i = \frac{\sum_{j=1}^{N_A} k_{i,j}}{\frac{1}{\tau_D} + \sum_{j=1}^{N_A} k_{i,j}} = \frac{\sum_{j=1}^{N_A} (R_0/R_{i,j})^6}{1 + \sum_{j=1}^{N_A} (R_0/R_{i,j})^6}, \tag{5}$$

where $R_{i,j}$ is the distance between the i^{th} donor and j^{th} acceptor.

The mean probability of energy transfer events for all donor molecules gives the energy transfer efficiency E for the entire system:

$$E = \langle p_i \rangle_{N_D}. \tag{6}$$

Steady-state FRET simulation

To analyze the experimental steady-state fluorescence data, steady-state FRET simulation is employed. The main advantage of this approach over Monte Carlo time-resolved simulation is its simplicity and high speed. The simulation starts with the generation of the structural model for the protein-lipid system. This model provides the coordinates of each donor and acceptor. The energy transfer efficiency E is then calculated using Eqs. 5 and 6. Because of the stochastic nature of the structural model, the resulting energy transfer efficiency contains stochastic deviations. Therefore the simulations are executed several times to make the results statistically relevant. The flow diagram of the simulation is shown in Fig. 3 and described below.

1. The parameters of the system are set (block *1*).
2. The structural model of a membrane with embedded proteins is created in accordance with the input parameters. The coordinates and orientation of the proteins provide information about the locations of donors and acceptors in the system (block *2*).
3. For each donor (denoted as i) the distances to all acceptors are considered and the probability of energy transfer (to any of them) is calculated using Eq. 5 (blocks *3–5*).
4. The mean probability of energy transfer among all donors results in the energy transfer efficiency for the whole system (Eq. 6).
5. Steps 2–4 (and blocks *2–6* in the flow diagram) are repeated for a number of times to decrease the effect of the randomness of the protein distribution. In our calculations we executed the simulation for 100 times.

Analytical model of FRET in planar systems

An analytical expression for FRET in a planar system was initially developed by Wolber and Hudson (12) and further enhanced by Davenport et al. (14). In these models, acceptors were considered as molecular systems of infinitesimal size uniformly distributed in a plane. The original equations by Davenport et al. can be modified to describe the energy transfer in the systems of M13 coat protein incorporated into lipid bilayers. The resulting analytical expression for the energy transfer efficiency E in the considered system is

$$E = 1 - \frac{1}{\tau_D}\int_0^\infty \rho_D(t) \times q_\sigma(t) \times \frac{r_{ul} + q_{intra}(t)}{1 + r_{ul}} dt, \tag{7}$$

where ρ_D is the fluorescence decay of a single donor, τ_D is the donor lifetime, and q_σ and q_{intra} are the quenching contributions of inter- and intramolecular energy transfer, respectively. The derivation of Eq. 7 and a further description of the expressions for ρ_D, q_σ, and q_{intra} are given in Appendix B.

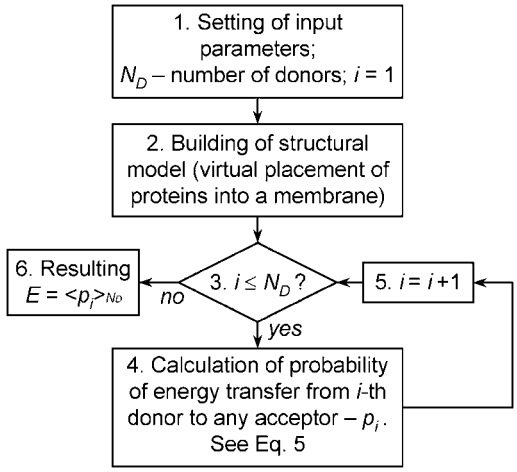

FIGURE 3 Flow diagram of a single simulation of energy transfer in a protein-lipid system.

Simulation-based fitting approach to experimental data analysis

The FRET model developed for M13 coat protein incorporated into lipid bilayers is used to analyze experimental data via the simulation-based fitting (SBF) approach. The scheme of SBF has been discussed in detail recently (18). As a measure of the goodness of the fit, the criterion was introduced

$$\chi^2 = \sum_{i=1}^{N} \left(E_i^e - E_i^s\right)^2, \quad (8)$$

where N is the number of data points, E_i^e the experimentally obtained energy transfer efficiency, and $1E_i^s$ the simulated energy transfer efficiency. To fit the modeled energy transfer efficiencies to the experimental ones, an optimization algorithm should be used. In our case, gradient optimization techniques are not applicable to fit the data, because of the stochastic behavior of the error function χ^2. Therefore, to perform a simultaneous fit of all experimental data, the Nelder-Mead simplex method (34) is used. This method provides a reasonable convergence and is not extremely time-consuming. To increase the robustness of the method and the precision of the solution, a global analysis approach is chosen, and therefore all experimental data were fitted simultaneously (35).

Because of the stochastic behavior of the FRET model, the error function χ^2 is stochastic as well, and the parameters obtained after each fit contain random deviations that are dependent on the sensitivity of the energy transfer to variations of the parameters. Therefore, to deal with this stochastic effect and to avoid possible local minima, the fitting procedure is performed 100 times with different initial estimations of the fitting parameters. The methodology used for the analysis of the resulting solutions and the selection of the representative solutions in terms of an optimal 20% elite subset is given in Appendix C.

All models were realized as C++ classes. The Borland C++ Builder 6.0 environment was used to combine the developed models, OpenGL visualization and SBF fitting algorithms into a software tool called FRETsim. The C++ classes and software are available from the authors upon request.

RESULTS

Experimental energy transfer efficiencies

An example of the experimentally obtained excitation spectra at different L/P ratios is presented in Fig. 4. The increase of the fluorescence intensity at the donor absorption wavelength (290 nm) clearly shows the increasing effect of energy transfer.

The mutants that were selected for our experiments (Y24C, G38C, and T46C) have their cysteines, and therefore the AEDANS labels, on the boundaries or close to the center of the transmembrane α-helix, which ranges from ~25–45 amino residues (25,28,30). For mutant G38C, two FRET titration experiments were performed at different values of r_{LP} (and also acceptor concentrations) to study its effect of protein association, given by parameter k. As a result of titration experiments on Y24C, T46C, and the double experiment on G38C mutants, four data series were obtained. The experimental L/P ratios r_{LP}, the unlabeled/labeled protein ratios r_{ul}, and resulting energy transfer efficiencies are presented in Table 1. The behavior of the energy transfer efficiency for different mutants as a function of r_{ul} is illustrated in Fig. 5.

Förster distance

The value of Förster distance R_0, needed for simulation of energy transfer, was calculated using Eq. 4. In this equation

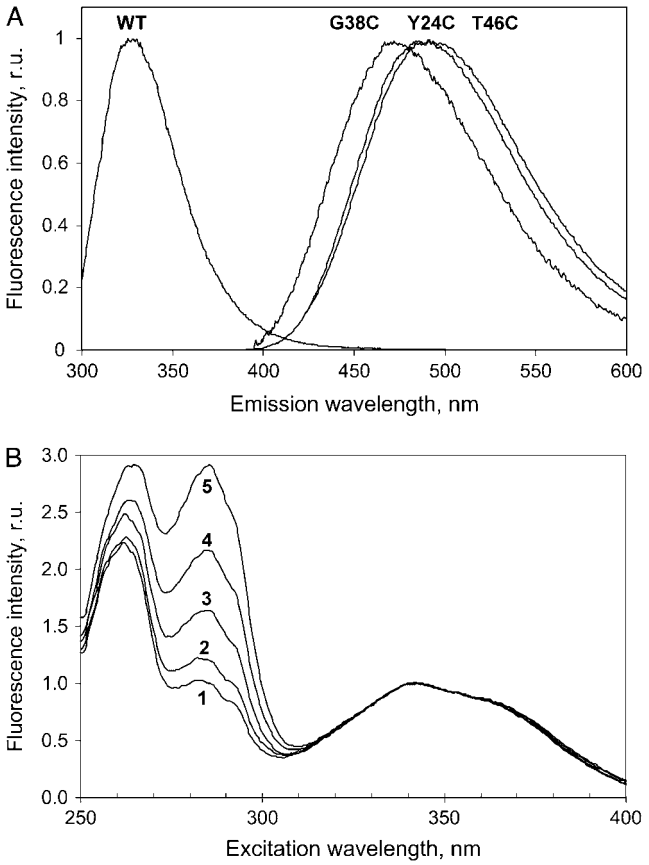

FIGURE 4 (*A*) Emission spectrum of wild-type proteins (*WT*) showing the Trp fluorescence, and emission spectra of mutant proteins Y24C, G38C, and T46C with AEDANS-labeled Cys at positions 24, 38, 46 after subtraction of the fluorescence of equimolar WT samples. Note that almost no Trp fluorescence can be observed at the AEDANS emission maxima. (*B*) Experimental excitation spectra obtained for mutant 38 at different titration points of wild-type proteins. The emission was detected at 470 nm. The labels *1–5* correspond to r_{ul} values of 6, 10, 19, 33, and 54, respectively. The lipid/protein ratios r_{LP} are 209, 128, 71, 45, and 28, respectively (see data set 2 in Table 1). The sample showing the highest peak at 290 nm (spectrum *5*) has the highest protein density (lower r_{LP}) and r_{ul}. Although the efficiency of energy transfer (Fig. 5) for this case is smallest, the overall energy absorbed by the donors in such a system, and therefore the transferred (intermolecular), is higher than for the other values of r_{LP} and r_{ul}.

$Q_D = 0.23$ was taken, which is the quantum yield of tryptophan in dimyristoyl phosphatidylcholine (DMPC) bilayers (36). The overlap integral J is calculated from the emission spectrum of the wild-type protein and the absorption spectrum of the AEDANS-labeled Y21A/Y24A/W26A/G23C mutant, which has no tryptophan at position 26. This results in a value of 5.96×10^{-15} M^{-1} cm^3. For small proteins and peptides, as is the case for M13 coat protein, the orientation factor κ^2 can be approximated by its isotropic dynamic average, giving a value of 2/3 (30,33,37–39). For simplicity the refractive index of the medium is assumed to be constant, and equal to 1.4 (5,14). These parameters result in a Förster radius R_0 of 24 Å. It should be noted that the excitation band of AEDANS, with its maximum at

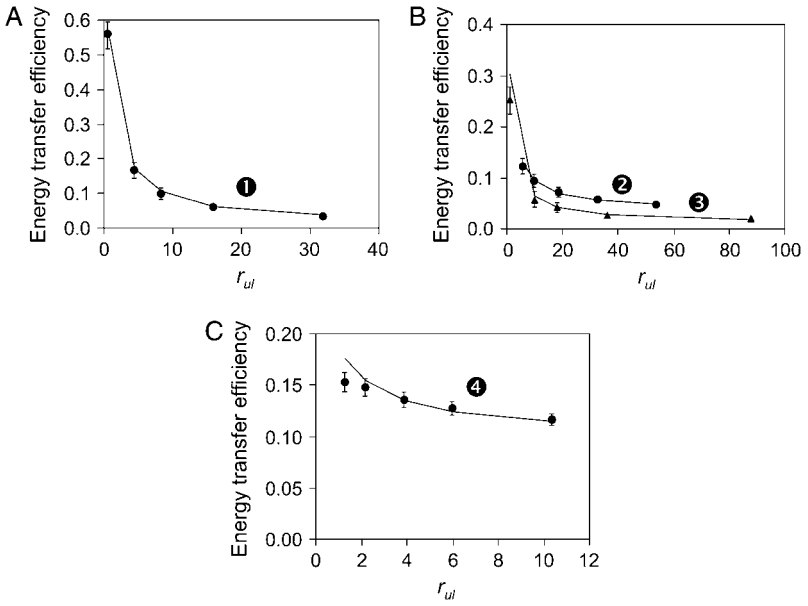

FIGURE 5 Experimental energy transfer efficiencies E (*solid dots* and *triangles*) and their approximation by the model (*solid line*) after global analysis versus the ratio between unlabeled and labeled proteins r_{ul}. (*A*) Mutant 24; (*B*) mutant 38; (*C*) mutant 46. The labels *1–4* refer to the corresponding data sets in Table 1. In panel *B*, the dots indicate data set 2 and the triangles data set 3. The error bars correspond to the maximal deviations of the data points observed during the experiments.

~340 nm, does not change with the position of the labeled cysteine. This implies that the Förster distance for the donor-acceptor pair is equal for all mutants.

Determination of bilayer topology of the protein

All four sets of experimental data were fitted simultaneously. The fitting procedure included 50 iterations of the Nelder-Mead simplex method. To avoid local minima, the fitting procedure was independently repeated 100 times with different initial estimations of the desired parameters: L, k, θ, ψ, and d. The values of initial estimations were randomly selected from the parameter ranges, presented in Table 2. The calculation of each single solution took ~20 min on a computer with a Pentium 4 processor (each simulation takes 1–5 s). Because the calculation of each solution is an independent task, the fitting was parallelized between several computers. The solutions found were analyzed as described in Simulation-Based Fitting Approach to Experimental Data Analysis. The resulting χ^2 for the elite set varies from 0.0039 to 0.0048, and the discarded solutions had a value χ^2 ranging from 0.0048 to 0.1.

The resulting values together with the standard deviations inside the elite set of solutions are presented in Table 3. This table also shows a compilation of the values known from the literature. The best fitting results are presented in Fig. 5 together with the experimental data.

DISCUSSION

Measuring strategy

In this study we aimed at the development of a methodology based on a combination of FRET spectroscopy and computer simulation, thereby providing information about the position and protein-protein associations in a membrane system. By assuming a helical structure for the fluorescent-labeled protein (or its domain), the proposed approach is able to determine both its topology and bilayer embedment in terms of protein tilt angle, direction of tilt, and protein depth in the membrane. Moreover, the method provides a quantitative analysis of the protein-protein associations, which can hardly be performed by other spectroscopic methods. In the case of a nondilute protein-lipid system with randomly distributed proteins, the energy of donor excitation can be transferred both intra- and intermolecularly. Because the aggregation behavior of M13 coat protein in lipid vesicles is not well documented, and cannot be excluded even at high L/P ratios, the efficiency of the intermolecular energy transfer component may partly arise from relatively short donor-to-acceptor distances in protein aggregates.

Being incorporated into the membrane, the proteins form two planes of donor and two planes of acceptor molecules, originating from parallel and antiparallel orientations of the proteins. The intermolecular energy transfer is influenced,

TABLE 3 Resulting parameters of the model for the protein-lipid system applied to the transmembrane domain of M13 major coat protein incorporated into DOPC/DOPG bilayers and the corresponding values known from the literature

Parameter	Value found	Previously reported value	Reference
L	0.28 ± 0.03	~0.2	Spruijt et al. (31)
k	0.03 ± 0.01	~0	Fernandes et al. (23)
θ	18 ± 2°	19 ± 1°	Koehorst et al. (25)
		26°	Marassi and Opella (27)
		20 ± 10°	Glaubitz et al. (26)
ψ	61 ± 7°	60°	Koehorst et al. (25)
D	8.5 ± 0.5 Å	8.9 Å	Koehorst et al. (25)

among other factors, by the distances between the donor and acceptor planes, which are determined by the z-coordinates of the fluorescent labels. Structural parameters describing the embedment and orientation of the protein, such as d, θ, and ψ (see Fig. 1), can change the positions of the planes, and therefore can be tracked by analyzing energy transfer processes.

The selection of mutants Y24C, G38C, and T46C was given by two rationales. First, the selected labeling sites should be located in an α-helical part of the M13 coat protein. This condition arises from the assumption of an α-helical protein model. Second, the selected sites should present maximally diverse intramolecular distances and acceptor positions inside the membrane, to increase the precision of the parameter determination. Therefore, sites should be located preferably at the edges of such a helical part. An α-helical conformation was suggested for positions from ~25 to 45 in the transmembrane domain (25,28,30). Therefore we selected the mutants Y24C, G38C, and T46C as labeling sites. To study possible effects of protein aggregation, additional experiments were performed for the G38C mutant at high and low L/P ratios (see Table 1).

For an ideal case of independent parameters and independent experiments without any distortion in the obtained data, the number of experiments N should be equal to n ($N = n$), where n is the number of unknown parameters. However, if the data set contains noise, the number of equations should be larger than the number of parameters (i.e., $N > n$). Obviously, the more data provided, the higher the precision one would get. In our specific situation, each of the data series (as shown in Table 1 and Fig. 5) can be considered as two independent points representing the intra- and intermolecular energy transfer. Therefore, for the situation of an α-helical protein model with five unknown structural parameters (giving $n = 5$, i.e., θ, d, ψ, L, and k), at least one independent data series coming from each of the three selected mutants is needed (giving $N = 6$). Of course, including additional series would enhance the precision of the determination of the parameters. Thus, as a rule of thumb, at least three donor-acceptor pairs would be needed that are regularly spread over the protein transmembrane domain.

To determine the energy transfer parameters, fluorescence excitation spectroscopy was used by monitoring the acceptor excitation at wavelengths 470–490 nm. Here the acceptor fluorescence was monitored, thereby optically selecting only the acceptor-labeled mutants, and discriminating between fluorescence resulting from donor-to-acceptor energy transfer and fluorescence resulting from direct excitation. There are two advantages of recording the acceptor fluorescence excitation over the donor fluorescence. First, the intensity of the background fluorescence between 450 and 550 nm is less than in the UV region (tryptophan/donor fluorescence is between 300 and 350 nm). Second, for an experiment in which the donor is monitored, varying the wild-type protein concentration or varying the mutant protein concentration would change the concentration of that donor, while in our approach the concentration of the monitored acceptors is kept constant. The small lifetime of tryptophan (~3.6 ns) allows us to assume that there is no lateral mobility in the system that can significantly change the donor-acceptor distance.

To separate intra- and intermolecular energy transfer contributions, we performed titration experiments in which mixtures of a fixed amount of labeled protein mutants and different amounts of wild-type protein were reconstituted into lipid vesicles. Both unlabeled mutant and wild-type protein can be considered spectroscopically identical as donor-containing molecules without acceptor label; however, labeled mutants contain both a donor and acceptor.

Validation of the simulation model

Before applying the protein-lipid model and the SBF approach to real experimental data, both the model and the approach should be validated. As a first step, the energy transfer efficiency is calculated for a system with different L/P ratios r_{LP} (for simplicity, we consider a constant $r_{ul} = 0$) and compared with results of the modified Davenport's analytical model, Eq. 7. The comparison is carried out for different values of r_{LP}, which influence the acceptor surface density. The resulting energy transfer efficiencies are plotted in Fig. 6, using a value for D_P, and consequently the exclusion distance in Davenport's model, of 10 Å, which is about the diameter of a transmembrane protein domain. The plot shows a deviation of the analytically obtained energy transfer efficiencies from the simulated ones. This finding provoked us to perform an additional study on the applicability of the analytical solution. As was mentioned before, the analytical solution is based on a number of simplifications; one of those is the assumption of an infinitely small acceptor size. To check this situation, a comparison is carried out by assuming a small transmembrane protein domain with an exclusion distance D_P of 1 Å. For such a system, a complete correspondence between the simulated and analytically calculated energy transfer efficiency is observed (Fig. 6).

From the comparison, it is clear that the steady-state simulation model of FRET gives the same results as the extended well-known analytical solution of Davenport et al. (14) (Eq. 7) in the case of small acceptor-labeled molecules. However, if the size of the molecules becomes comparable to the Förster distance, the simulation-based approach should be used rather than the analytical model. It is clear that the limiting situation for small molecular sizes of the simulation-based approach corresponds to the analytical solution. The simulation-based approach is more general and powerful than the analytical model, and can be applied for the analysis of donor-acceptor systems with any geometry.

Testing of the simulation-based fitting approach

A numerical test was performed to prove the applicability of the SBF approach to the problem of M13 coat protein

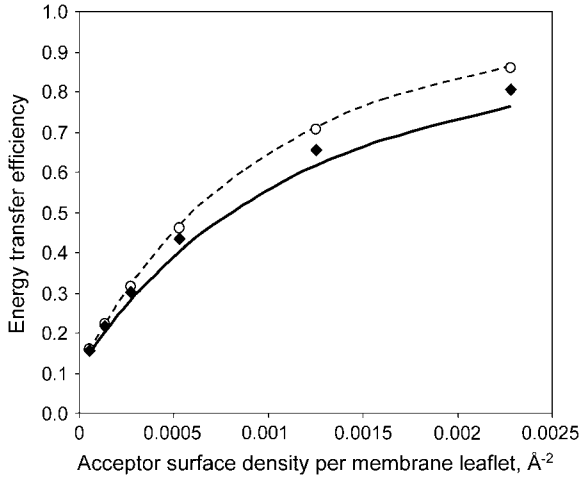

FIGURE 6 Comparison of simulation results with analytical solutions for different sizes of proteins. (*Solid line*) Analytical result for $D_P = 10$ Å. (♦) Simulation results with the same protein exclusion distance $D_P = 10$ Å. (*dotted line*) Analytical energy transfer efficiency. (○) Simulated energy transfer efficiency for $D_P = 1$ Å. All calculations were performed with the following protein parameters: $n_A = 46$, $n_D = 26$, $\theta = 16°$, $\psi = 50°$, $d = 10$ Å, $S_L = 72$ Å^2, and $R_0 = 24$ Å. The corresponding parameters for the analytical model are $h_I = 26.8$ Å, $h_{II} = 13.1$ Å, and $R_{intra} = 33.9$ Å.

structure determination and to find the optimal elite subset size. In this test, synthetic FRET data were generated using the model with values of parameters close to those determined experimentally for M13 coat protein. The simulation was performed for 1000 proteins and the results were averaged for 1000 simulations. This provided us with synthetic data containing a very small randomness. Then these synthetic data were analyzed via the SBF approach as mentioned before, and the solutions were handled as shown in Appendix C. The smallest deviation from the original values of the parameters was found for a 20% elite subset. The results are presented in Table 4. The random spread of the solutions inside the elite set was close to that obtained during the analysis of the experimental data (Table 3).

The values found for the association coefficient k and lipid loss L (see Table 4) are very close to the original ones, and have a relatively small error. These two parameters influence the surface density of the label and therefore have a strong effect on the intermolecular FRET. Despite some correlation between k and L, the method is able to determine both parameters quite well. From the results in Table 4, it follows that the values of the protein depth d and protein tilt angle θ are close to the original ones. The direction of protein tilting ψ has a substantial large variation. The reason for the spread in ψ is that this parameter does not significantly influence the position of the donor and acceptor planes.

To study the possible effect of experimental noise on the resulting data, we introduced Gaussian noise to the synthetic data, and performed a number of fittings. Each SBF was performed with its own random deviations in the data. The standard deviation of each data point was calculated according to maximal deviations observed in the FRET experiments, which are ~10% for points with a low r_{ul} and ~5% for points with a high r_{ul} (see the *error bars* in Fig. 5). The results of the fitting of the noisy data are given Table 4. As can be seen, the parameters L, k, and d are only slightly affected by introducing noise, indicating that they can be determined quite precisely from the FRET experiments. This stability to noise can be explained by the fact that we use a global analysis approach and that for each mutant we have five data points. The angular parameters (θ and ψ) tend to deviate from the original value, indicating that they are relatively more sensitive to noise in the experimental data.

TABLE 4 Original and calculated values of the model parameters after analysis of synthetic FRET data by means of an SBF approach

Parameter	Original value in synthetic data simulation	Value found after SBF analysis with no noise added to synthetic data	Value found after SBF analysis with additional noise in synthetic data
L	0.3	0.29 ± 0.01	0.30 ± 0.03
k	0.05	0.05 ± 0.01	0.07 ± 0.02
θ	20°	19 ± 3°	16 ± 3°
ψ	60°	61 ± 8°	47 ± 15°
D	9 Å	8.9 ± 0.2 Å	9.2 ± 1 Å

To introduce noise in the synthetic data, a standard deviation of ~10% is used for points with a low r_{ul} and ~5% for points with a high r_{ul} (see the *error bars* in Fig. 5).

From this test, it can be concluded that the application of the described biophysical model together with the SBF approach to data analysis can determine the protein location in a bilayer, and the protein-protein association. This result gives us the confidence to apply the methodology to analyze our experimental FRET data.

Parameters determined

Table 3 summarizes the resulting parameters and the corresponding values known from the literature. Variation of the parameters within the error limits given in Table 3 does not result in values of χ^2 higher than 0.0048 (in fact, all acceptable solutions have χ^2 values between 0.0039 and 0.0048, see Fig. 7 *A*). For example, increasing θ by 20° to 38° increases χ^2 to 0.0113. This χ^2 value is far above the limit of 0.0048 that was taken as acceptable.

The actual value of the parameter describing the lipid loss during dialysis L is unknown and has to be determined using the SBF approach from our experimental data. The value found is 0.28, which means that ~28% of the lipids are washed-out from the sample into the buffer during dialysis. This value is in reasonable agreement with the lipid loss of 20% as estimated from biochemical analysis (31). All our experiments were performed under identical conditions and using the same protocol. This allows us to assume that the

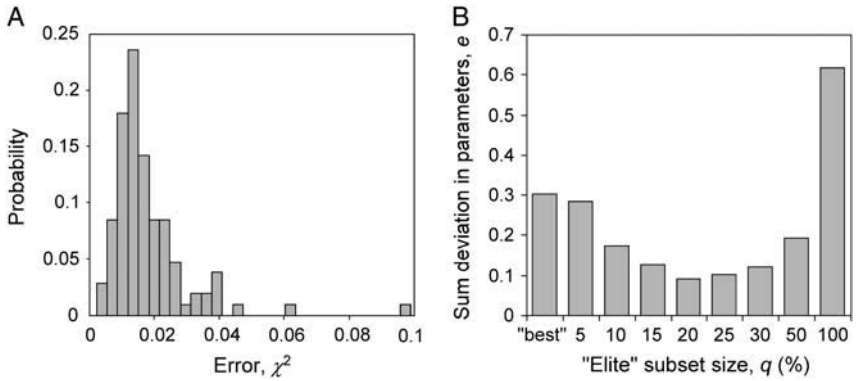

FIGURE 7 (*A*) Distribution of χ^2 of the solutions found after 100 runs of SBF on experimental data and (*B*) behavior of the sum parameter deviation e (Eq. C1) with respect to the elite subset size q. The result in panel *B* is obtained after averaging the results of three independent numerical simulations. For all of them, the optimal q was ~20%.

lipid loss is constant for all experiments. The small value of the association constant k indicates that the proteins have no tendency to aggregate under the experimental conditions. This is in agreement with earlier observations of the protein in DOPC/DOPG mixtures (23). Again, some correlation between parameters k and L was found. This effect is included in the uncertainty limits for the parameters in Table 3.

The resulting protein depth d of 8.5 Å is very close to the value of 8.9 Å as found from fluorescent experiments in DOPC/DOPG (25). The tilt angle of the transmembrane helix $\theta = 18°$ is somewhat smaller than the value $\theta = 26°$ arising from solid-state NMR (27). However, it is within the range of $20 \pm 10°$, as found earlier from solid-state ^{13}C NMR (26). From Stokes-shift experiments, a range of tilt angles from 18 to 28° was estimated (25). In this work (25), the tilt angle is given as a function of the distance between the AEDANS moiety and the α-helix axis. A tilt angle of $19 \pm 1°$ corresponding to the distance $l_A = 9.5$ Å, used in our work, is in excellent agreement with our value of $18 \pm 2°$. The direction of the protein tilt ψ is the least-sensitive parameter in our case. Nevertheless, our value of $61 \pm 7°$ is close to 60°, as found previously (25). This comparison shows that our model is performing well, certainly by taking into account that only three different mutants were used.

From Fig. 5, it can be noticed that some fits are not ideal. The reason for these deviations between simulated and experimental efficiencies could be related to the fact that the long AEDANS label arm is mobile within a restricted space angle, the size and direction of which differs for different mutants (30). A future enhancement of the model could be the implementation of the entire AEDANS conformational space for each mutant instead of assuming a constant acceptor arm normal to the helix axis. A further improvement of the precision of our model can be achieved by using the fluorescent data of the AEDANS (25) in a general global optimization algorithm. We are currently working on these challenging ideas.

The methodology developed here is not limited to M13 major coat protein and can be used, in principle, to study the bilayer embedment and structure of any α-helical single transmembrane protein (or peptide), and with some adaptations to transmembrane domains of larger membrane proteins. For example, the method was successfully applied to study the aggregation of various WALP peptides in lipid bilayers of different thickness (40).

APPENDIX A: DERIVATION OF ENERGY TRANSFER EFFICIENCY

Consider a protein-lipid system containing two types of fluorescently labeled proteins—with a single donor (denote its quantity by C_u) and with a donor and acceptor (denote the quantity by C_l). Let us introduce two efficiencies of energy transfer E_u and E_l, characterizing energy transfer for the first and second protein population. The total energy transfer efficiency E for the system is then given by

$$E = \frac{C_u}{C_u + C_l} E_u + \frac{C_l}{C_u + C_l} E_l. \tag{A1}$$

Consider now the acceptor excitation spectrum for such a system in a general case of

$$F^\lambda = \left(\Theta^\lambda_A + \Theta^\lambda_u + \Theta^\lambda_l\right)\gamma, \tag{A2}$$

where Θ_A is the direct acceptor excitation at wavelength λ, $\Theta^\lambda{}_u$ is the excitation due to energy transfer from unlabeled proteins, $\Theta^\lambda{}_l$ is the excitation caused by energy transfer from labeled proteins (both intra- and intermolecular), and γ is a constant that depends on the apparatus and experimental conditions. Taking into account the extinction coefficients of donor and acceptor and protein quantities, this equation can be rewritten in the following form

$$F^\lambda = \gamma\left(C_l\varepsilon^\lambda_A + C_u\varepsilon^\lambda_D E_u + C_l\varepsilon^\lambda_D E_l\right), \tag{A3}$$

where ε^λ_u is the extinction coefficient of acceptors at wavelength λ, and ε^λ_D is the extinction coefficient of the donors. At $\lambda = 290$ nm, the extinction coefficients are nonzero both for our donor (Trp-26) and acceptor (AEDANS). However, at $\lambda = 340$ nm, $\varepsilon^{340}_D = 0$. Taking into account the fluorescence at these two wavelengths and expressing the partial efficiencies via Eq. A1, the following descriptions for the fluorescence of the protein-lipid system can be obtained:

$$F^{290} = \gamma\left(C_l\varepsilon^{290}_A + \varepsilon^{290}_D(C_u + C_l)E\right), \tag{A4}$$

$$F^{340} = \gamma C_l\varepsilon^{340}_A. \tag{A5}$$

Dividing Eq. A4 by A5 and making simple rearrangements the following equation is obtained:

$$\left(\frac{F^{290}}{F^{340}}-\frac{\varepsilon_A^{290}}{\varepsilon_A^{340}}\right)\frac{\varepsilon_A^{340}}{\varepsilon_D^{290}}=\left(1+\frac{C_u}{C_l}\right)E. \tag{A6}$$

By introducing the ratio of the number of unlabeled to labeled proteins, r_{ul}, Eq. A6 can then be presented in the form

$$E=\frac{1}{1+r_{ul}}\left(\frac{F^{290}}{F^{340}}-\frac{\varepsilon_A^{290}}{\varepsilon_A^{340}}\right)\frac{\varepsilon_A^{340}}{\varepsilon_D^{290}}. \tag{A7}$$

APPENDIX B: ANALYTICAL EQUATION FOR FRET IN SYSTEMS OF M13 COAT PROTEIN PROTEINS INCORPORATED INTO A LIPID BILAYER

Consider a system of labeled and unlabeled M13 coat protein incorporated into a lipid bilayer. Let r_{ul} be the molar ratio of the labeled and unlabeled proteins. The time decay of the fluorescence intensity, $\rho(t)$, of the donor in this system can then be described by

$$\rho(t)=\frac{r_{ul}}{1+r_{ul}}\rho_u(t)+\frac{1}{1+r_{ul}}\rho_l(t), \tag{B1}$$

where ρ_u is the fluorescence decay of the unlabeled proteins, and ρ_l the fluorescence decay of the labeled proteins. The coefficients in front of $\rho_{u,l}$ in Eq. B1 are the fractions of labeled and unlabeled proteins expressed in terms of r_{ul}.

The fluorescence decay of donors attached to unlabeled proteins ρ_u is affected by acceptors of other proteins, distributed around. For labeled proteins, the intramolecular energy transfer should be taken into account as well. Thus,

$$\rho_u(t)=\rho_D(t)\times q_\sigma(t,R_0,\sigma,h,D_P), \tag{B2}$$

$$\rho_l(t)=\rho_D(t)\times q_\sigma(t,R_0,\sigma,h,D_P)\times q_{intra}(t), \tag{B3}$$

where ρ_D is the fluorescence of a single donor, q_σ the quenching effect by distributed acceptors, and q_{intra} the quenching effect by intramolecular energy transfer in labeled proteins. We assume now that the donor fluorescence has a single lifetime and can be described by

$$\rho_D(t)=\exp(-t/\tau_D), \tag{B4}$$

where τ_D is a single donor lifetime. Alternatively, all expressions presented below may easily be reproduced for multiexponential donor fluorescence (41).

The quenching by intramolecular energy transfer is given by

$$q_{intra}(t)=\exp\left(-t(R_0/R_{intra})^6/\tau_D\right), \tag{B5}$$

where R_{intra} is the intramolecular donor-acceptor distance. Consider now the quenching due to distributed acceptors. The overall surface density of acceptors is given by

$$\sigma=\frac{N_A}{S}=\frac{N_l}{S_L N_L/2+S_P(N_l+N_u)}, \tag{B6}$$

where S is the area of the entire membrane, N_A the number of acceptors in the system, N_l the number of labeled proteins, N_u the number of unlabeled proteins, N_L the number of lipids, S_L the area occupied by a single lipid molecule, and S_P the area occupied by a single protein molecule. Taking into account the definitions of r_{LP} and r_{ul}, and considering cylindrical proteins, Eq. B6 can be presented in the form

$$\begin{aligned}\sigma&=\left[(S_L r_{LP}/2+S_P)(1+r_{ul})\right]^{-1}\\&=2\left[(S_L r_{LP}+D_P^2\pi/2)(1+r_{ul})\right]^{-1}.\end{aligned} \tag{B7}$$

Because of the possibility of parallel and antiparallel protein orientations, the initial acceptor density σ is divided over the two leaflets. For each leaflet, the acceptor density σ_1 is given by

$$\sigma_1=\sigma/2=\left[\left(S_L r_{LP}+D_P^2\pi/2\right)(1+r_{ul})\right]^{-1}. \tag{B8}$$

Donors are divided over the two leaflets as well. The symmetry of the system then leads to an equivalence of relative distances between each donor plane and two acceptor planes. Therefore, the system can be substituted with a system containing one layer of donors and two layers of acceptors at the distances of

$$h_I=|Z_D-Z_A|,\ \text{and}\ h_{II}=|Z_D+Z_A|, \tag{B9}$$

where Z_D and Z_A are the z coordinates (in the membrane axis system) of a donor and acceptor, respectively, attached to a protein with an upright orientation.

The analytical solution for the donor fluorescence decay in the presence of uniformly distributed acceptors in a plane was given by Davenport et al. (14). Taking into account two layers of acceptors located at h_I and h_{II}, the quenching effect on the donor fluorescence is given as

$$\begin{aligned}q_\sigma(t)&=q_\sigma^I(t)\times q_\sigma^{II}(t)=\\&\exp\left\{-2\pi\sigma_1 h_I^2\int_{\alpha=0}^{\frac{h_I}{\sqrt{h_I^2+D_P^2}}}\left(1-\exp\left[-\frac{t}{\tau_D}\left(\frac{R_0\alpha}{h_I}\right)^6\right]\right)\alpha^{-3}d\alpha\right.\\&\left.-2\pi\sigma_1 h_{II}^2\int_{\beta=0}^{\frac{h_{II}}{\sqrt{h_{II}^2+D_P^2}}}\left(1-\exp\left[-\frac{t}{\tau_D}\left(\frac{R_0\beta}{h_{II}}\right)^6\right]\right)\beta^{-3}d\beta\right\}.\end{aligned} \tag{B10}$$

The energy transfer efficiency can be calculated using the relative integrated fluorescence intensity of the donors in the presence and absence of acceptors as

$$E=1-\int_0^\infty\rho(t)dt\Big/\int_0^\infty\rho_D(t)dt. \tag{B11}$$

The integrated fluorescence of a single donor in the case of one exponential decay equals to τ_D. After substitution of ρ, the energy transfer efficiency E can be expressed in terms of ρ_D, q_σ, and q_{intra},

$$E=1-\frac{1}{\tau_D}\int_0^\infty\rho_D(t)\times q_\sigma(t)\times\frac{r_{ul}+q_{intra}(t)}{1+r_{ul}}dt. \tag{B12}$$

APPENDIX C: ANALYSIS OF THE SOLUTIONS OBTAINED BY SBF

The FRET model that is used in our SBF fitting has a random nature and therefore the error function χ^2 (Eq. 8) is a stochastic one. To deal with this stochastic effect, the fitting procedure needs to be performed several times (we take 100, which is found to be sufficiently large) with different starting fitting parameters. This approach results in a distribution of solutions and each of the resulting solutions has a different χ^2 value. A typical distribution of resulting χ^2 values is shown in Fig. 7 *A*.

In this case, the selection of the parameter set corresponding to the minimal χ^2 is not statistically correct, because a low χ^2 can be the result of a random deviation. At the same time, averaging of all solutions found will lead to an incorrect result as well, because many solutions with a high χ^2 are included. These solutions do not show a reasonable fit between the modeled and experimental data and appear only because the optimization algorithm is falling into false local minima.

To reduce the randomness of a single solution and to find the best solution in the parameter space, we use the following approach, which is often found in evolutionary computing (42). A part of the solutions with the lowest χ^2 values is selected. This corresponds to selecting the quantile χ^2_q of the χ^2 distribution. The solutions with χ^2 less than the selected quantile are considered as an elite subset, and the mean value of the parameters inside this elite subset is then taken as the result of the fitting. The problem of this approach now reduces to finding the optimal size for the elite solutions, i.e., the value q in the quantile χ^2_q.

The selection of the optimal q is a problem-related task and cannot be analytically solved in general. Therefore, we employ an empirical approach. Using our numerical model, the analogs of experimental data were simulated for a known parameter vector $\boldsymbol{P}$, and these synthetic data were fitted by the same model. The resulting solutions were analyzed using the quantile approach with various values of q. This provides the resulting parameter vector $\boldsymbol{P}^*$. To validate the precision of the representative solutions found, we introduce a function e, which is the sum of the parameter deviations,

$$e = \sum_{i=1}^{n_p} \left(\frac{P_i^* - P_i}{P_i} \right)^2, \qquad \text{(C1)}$$

where n_p is the number of parameters, and P_i is i^{th} parameter from the parameter vector $\boldsymbol{P}$. The sum parameter deviation e is related to the inaccuracy in the resulting parameters. The behavior of the function e with respect to q for our FRET model is depicted in Fig. 7 *B*. On increasing q the error is decreasing, as would be expected, since the noise is reduced. However, after taking more solutions into account, the error is increasing again, because bad solutions are coming in. The minimal deviation in the parameters is reached for $q = 20\%$.

To be fully applicable, the algorithm needs all elite solutions belonging to the neighborhood of a single χ^2 minimum, and their differences should be caused by simulation randomness. If the solutions would form several separated clusters, the same approach should be applied to each of those clusters, and the solutions found should be considered as possible states for the system. However, this does not happen in our case. The additional advantage of the proposed algorithm is that it gives direct insight in the error range. The standard deviation of parameters inside the elite subset of solutions therefore can be used as a characteristic of the error range of the resulting solution.

We thank Ruud B. Spruijt for the preparation of the protein mutants and helpful comments on the work.

This work was supported by contract No. QLG-CT-2000-01801 of the European Commission (MIVase—New Therapeutic Approaches to Osteoporosis: Targeting the Osteoclast V-ATPase).

REFERENCES

1. Byrne, B., and S. Iwata. 2002. Membrane protein complexes. *Curr. Opin. Struct. Biol.* 12:239–243.
2. Arora, A., and L. K. Tamm. 2001. Biophysical approaches to membrane protein structure determination. *Curr. Opin. Struct. Biol.* 11: 540–547.
3. Torres, J., T. J. Stevens, and M. Samso. 2003. Membrane proteins: the ''Wild West'' of structural biology. *Trends Biochem. Sci.* 28:137–144.
4. Dos Remedios, C. G., and P. D. J. Moens. 1995. Fluorescence resonance energy transfer spectroscopy is a reliable ruler for measuring structural changes in proteins. Dispelling the problem of the unknown orientation factor. *J. Struct. Biol.* 115:175–185.
5. Lakowicz, J. R. 1999. Principles of Fluorescence Spectroscopy. Kluwer Academic/Plenum Publishers, New York.
6. Förster, T. 1965. Delocalized excitation and energy transfer. *In* Modern Quantum Chemistry. O. Sinanoglu, editor. Academic Press, New York.
7. Stryer, L. 1978. Fluorescence energy transfer as a spectroscopic ruler. *Annu. Rev. Biochem.* 47:819–846.
8. Lakey, J. H., D. Duche, J. M. Gonzalez-Manas, D. Baty, and F. Pattus. 1993. Fluorescence energy transfer distance measurements. The hydrophobic helical hairpin of colicin A in the membrane bound state. *J. Mol. Biol.* 230:1055–1067.
9. Li, M., L. G. Reddy, R. Bennett, N. D. Silva, L. R. Jones, and D. D. Thomas. 1999. A fluorescence energy transfer method for analyzing protein oligomeric structure: application to phospholamban. *Biophys. J.* 76:2587–2599.
10. Förster, T. 1948. Intermolecular energy migration and fluorescence. *Ann. Phys.* 2:55–75.
11. Fleming, P. J., D. E. Koppel, A. L. Y. Lau, and P. Strittmatter. 1979. Intramembrane position of the fluorescent tryptophanyl residue in membrane-bound Cytochrome b5. *Biochemistry.* 24:5458–5464.
12. Wolber, P. K., and B. S. Hudson. 1979. An analytic solution to the Förster energy transfer problem in two dimensions. *Biophys. J.* 28:197–210.
13. Dewey, T. G., and G. G. Hammes. 1980. Calculation on fluorescence resonance energy transfer on surfaces. *Biophys. J.* 32:1023–1035.
14. Davenport, L., R. E. Dale, H. R. Bisby, and R. B. Cundall. 1985. Transverse location of the fluorescent probe 1,6-diphenyl-1,3,5-hexatriene in model lipid bilayer membrane systems by resonance excitation energy transfer. *Biochemistry.* 24:4097–4108.
15. Frederix, P. L. T. M., E. L. de Beer, W. Hamelink, and H. C. Gerritsen. 2002. Dynamic Monte Carlo simulations to model FRET and photobleaching in systems with multiple donor-acceptor interactions. *J. Phys. Chem. B.* 106:6793–6801.
16. Berney, C., and G. Danuser. 2003. FRET or no FRET: a quantitative comparison. *Biophys. J.* 84:3992–4010.
17. Yatskou, M. M., H. Donker, R. B. M. Koehorst, A. van Hoek, and T. J. Schaafsma. 2001. Energy transfer processes in zinc-porphyrin films studied by time-resolved fluorescence spectroscopy and Monte Carlo simulations. *Chem. Phys. Lett.* 345:141–150.
18. Nazarov, P. V., V. V. Apanasovich, V. M. Lutkovski, M. M. Yatskou, R. B. M. Koehorst, and M. A. Hemminga. 2004. Artificial neural network modification of simulation-based fitting: application to a protein-lipid system. *J. Chem. Inf. Comput. Sci.* 44:568–574.
19. Stopar, D., R. B. Spruijt, C. J. A. M. Wolfs, and M. A. Hemminga. 2003. Protein-lipid interactions of bacteriophage M13 major coat protein. *Biochim. Biophys. Acta.* 1611:5–15.
20. Spruijt, R. B., A. B. Meijer, C. J. A. M. Wolfs, and M. A. Hemminga. 2000. Localization and rearrangement modulation of the N-terminal arm of the membrane-bound major coat protein of bacteriophage M13. *Biochim. Biophys. Acta.* 1509:311–323.
21. Meijer, A. B., R. B. Spruijt, C. J. A. M. Wolfs, and M. A. Hemminga. 2001. Configurations of the N-terminal amphipathic domain of the membrane-bound M13 major coat protein. *Biochemistry.* 40:5081–5086.
22. Stopar, D., R. B. Spruijt, C. J. A. M. Wolfs, and M. A. Hemminga. 2002. Structural characterization of bacteriophage M13 solubilization by amphiphiles. *Biochim. Biophys. Acta.* 1594:54–63.
23. Fernandes, F., L. M. S. Loura, R. B. M. Koehorst, R. B. Spruijt, M. A. Hemminga, and M. Prieto. 2004. Quantification of protein-lipid selectivity using FRET: application to the M13 major coat protein. *Biophys. J.* 87:344–352.
24. Fernandes, F., L. M. S. Loura, M. Prieto, R. B. M. Koehorst, R. B. Spruijt, and M. A. Hemminga. 2003. Dependence of M13 major coat protein oligomerization and lateral segregation on bilayer composition. *Biophys. J.* 85:2430–2441.
25. Koehorst, R. B. M., R. B. Spruijt, F. J. Vergeldt, and M. A. Hemminga. 2004. Lipid bilayer topology of the transmembrane α-helix of M13 major coat protein and bilayer polarity profile by site-directed fluorescence spectroscopy. *Biophys. J.* 87:1445–1455.
26. Glaubitz, C., G. Grobner, and A. Watts. 2000. Structural and orientational information of the membrane embedded M13 coat protein by ^{13}C-MAS NMR spectroscopy. *Biochim. Biophys. Acta.* 1463:151–161.
27. Marassi, F. M., and S. J. Opella. 2003. Simultaneous assignment and structure determination of a membrane protein from NMR orientational restrains. *Protein Sci.* 12:403–411.

28. Papavoine, C. H. M., M. L. Remerowski, L. M. Horstink, R. N. H. Konings, C. W. Hilbers, and F. J. M. van de Ven. 1997. Backbone dynamics of the major coat protein of bacteriophage M13 in detergent micelles by ^{15}N nuclear magnetic resonance relaxation measurements using the model-free approach and reduced spectral density mapping. *Biochemistry*. 36:4015–4026.

29. Papavoine, C. H. M., B. E. C. Christiaans, R. H. A. Folmer, R. N. H. Konings, and C. W. Hilbers. 1998. Solution structure of the M13 major coat protein in detergent micelles: a basis for a model of phage assembly involving specific residues. *J. Mol. Biol*. 282:401–419.

30. Vos, W. L., R. B. M. Koehorst, R. B. Spruijt, and M. A. Hemminga. 2005. Membrane-bound conformation of M13 major coat protein: a structure validation through FRET-derived constraints. *J. Biol. Chem.* 280:38522–38527.

31. Spruijt, R. B., C. J. A. M. Wolfs, and M. A. Hemminga. 1989. Aggregation-related conformational change of the membrane-associated coat protein of bacteriophage M13. *Biochemistry*. 28:9158–9165.

32. Spruijt, R. B., C. J. A. M. Wolfs, J. W. G. Verver, and M. A. Hemminga. 1996. Accessibility and environment probing using cysteine residues introduced along the putative transmembrane domain of the major coat protein of bacteriophage M13. *Biochemistry*. 35:10383–10391.

33. Gustiananda, M., J. R. Liggins, P. L. Cummins, and J. E. Gready. 2004. Conformation of prion protein repeat peptides probed by FRET measurements and molecular dynamics simulations. *Biophys. J*. 86:2467–2483.

34. Nelder, J. A., and R. Mead. 1965. A simplex method for function minimization. *Comput. J*. 7:308–313.

35. Beechem, J. M., and L. Brand. 1986. Global analysis of fluorescence decay: applications to some unusual experimental and theoretical studies. *Photochem. Photobiol*. 44:323–329.

36. Fisher, C. A., and R. O. Ryan. 1999. Lipid binding-induced conformational changes in the N-terminal domain of human apolipoprotein E. *J. Lipid Res*. 40:93–99.

37. Loura, L. M. S., A. Fedorov, and M. Prieto. 1996. Resonance energy transfer in a model system of membranes: application to gel and liquid crystalline phases. *Biophys. J*. 71:1823–1836.

38. Lakshmikanth, G. S., K. Sridevi, G. Krishnamoorthy, and J. B. Udgaonkar. 2001. Structure is lost incrementally during the unfolding of barstar. *Nat. Struct. Biol*. 8:799–804.

39. Kamal, J. K. A., and D. V. Behere. 2002. Spectroscopic studies on human serum albumin and methemalbumin: optical, steady-state, and picosecond time-resolved fluorescence studies, and kinetics of substrate oxidation by methemalbumin. *J. Biol. Inorg. Chem*. 7:273–283.

40. Sparr, E., W. L. Ash, P. V. Nazarov, D. T. S. Rijkers, M. A. Hemminga, D. P. Tieleman, and J. A. Killian. 2005. Self-association of transmembrane α-helices in model membranes: importance of helix orientation and role of hydrophobic mismatch. *J. Biol. Chem.* 280:39324–39331.

41. Loura, L. M. S., A. Fedorov, and M. Prieto. 2001. Fluid-fluid membrane microheterogeneity: a fluorescence resonance energy transfer study. *Biophys. J*. 80:776–788.

42. Štrancar, J., T. Koklic, Z. Arsov, B. Filipic, D. Stopar, and M. A. Hemminga. 2005. Spin label EPR-based characterization of biosystem complexity. *J. Chem. Inf. Model*. 45:394–406.